IN THE CLAIMS

Please add new claim 57 and amend claims 1 and 17 as follows.

1. (currently amended)   A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

i) encoding two or more related biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a data structure to populate a data structure of product strings; and

v) determining [whether the product strings have] sequence identities of the product strings relative to [of greater than 30% with] at least one initial character string; and

vi) selecting [for production] one or more product biological molecules for production, wherein the one or more product biological molecules correspond [corresponding] to one or more of the product strings [determined to have] having greater than 30% sequence identity with the at least one initial character string.

2. (previously amended) The method of claim 1, wherein said encoding comprises encoding two or more nucleic acid sequences into said character strings.

3. (previously amended) The method of claim 2, wherein said two or more nucleic acid sequences comprise a nucleic acid sequence encoding a naturally occurring protein.

4. (previously amended) The method of claim 1, wherein said encoding comprises encoding two or more amino acid sequences into said character strings.

5. (previously amended) The method of claim 4, wherein said one or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

4

6. (original) The method of claim 1, wherein said biological molecules have at least 30% sequence identity with each other.

7. (original) The method of claim 1, wherein said selecting in (ii) comprises selecting substrings such that the ends of said substrings occur in string regions of about 3 to about 20 characters that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the same two strings.

8. (original) The method of claim 1, wherein said selecting in (ii) comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

9. (original) The method of claim 1, wherein said selecting in (ii) and concatenating comprises concatenating substrings from two different initial strings such that the concatenation occurs in a region of about three to about twenty characters having higher sequence identity between said two different initial strings than the overall sequence identity between said two different initial strings.

10. (previously amended) The method of claim 1, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

11 (canceled)

12 (previously amended). The method of claim 1, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

13. (original) The method of claim 11, wherein said method further comprises randomly selecting and altering one or more occurrences of a particular preselected character in said character strings.

14. (previously amended) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on an internet site.

15. (previously amended) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on a server.

5

15
16. (previously amended) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on a client linked to a network.

16 17. (currently amended)   A computer program product on a computer readable media comprising computer code that:

i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;

ii) selects at least two initial substrings from said character strings;

iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adds the product strings to a data structure to populate a data structure of product strings; and

v) determines [whether the product strings have] sequence identities [of greater than 30% with] of the product strings relative to at least one initial character string; and

vi) selects [for production of] one or more product biological molecules [corresponding] for production, wherein the one or more product biological molecules correspond to one or more of the product strings [determined to have] having greater than 30% sequence identity with the at least one initial character string.

17
18. (previously amended) The computer program product of claim 17, wherein said two or more biological molecules are nucleic acid sequences.

18
19. (previously amended) The computer program product of claim 17, wherein said two or more biological molecules are nucleic acid sequences of naturally occurring proteins.

19
20. (previously amended) The computer program product of claim 17, wherein said two or more biological molecules are amino acid sequences.

20
21. (previously amended) The computer program product of claim 17, wherein said biological molecules have at least 30% sequence identity.

21
22. (previously amended) The computer program product of claim 17, wherein said computer code selects substrings such that the ends of said substrings occur in string regions of about three to about twenty characters that have higher sequence identity with a corresponding

6

region of another of said initial character strings than the overall sequence identity between the two substrings.

22
23. (previously amended) The computer program product of claim 17, [16] wherein said computer code selects substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

23
24. (previously amended) The computer program product of claim 17, [16] wherein said computer code selects and concatenates substrings from two different initial strings such that the concatenation occurs in a region of about three to about twenty characters having higher sequence identity between said two different initial strings than the overall sequence identity between said two different initial strings.

24
25. (previously amended) The computer program product of claim 17, [16] wherein the computer code selects substrings by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the character strings, and selecting a character that is a member of an aligned pair for the end of one substring.

26. (canceled)

25
27. (previously amended) The computer program product of claim 17, [16] wherein said computer code additionally randomly alters one or more characters of said character strings.

26
28. (previously amended) The computer program product of claim 27, [25] wherein said computer code additionally randomly selects and alters one or more occurrences of a particular preselected character in said character strings.

29. (previously amended) The computer program product of claim 17, wherein said computer code is stored on media selected from the group consisting of magnetic media, optical media, optomagnetic media.

28
30. (previously amended) The computer program product of claim 17, [16] wherein said computer code is in dynamic or static memory of a computer.

31-44. (withdrawn) cancelled

29
45. (previously added) The method of claim 1, wherein the initial character strings of (i) are related.

7

30
46. (previously added) The method of claim 1, further comprising physically screening the molecule(s) represented by the product strings for one or more desired properties.

31
47. (previously added) The method of claim 1, further comprising determining a computationally predicted property for molecules represented by the product strings.

32
48. (previously added) The method of claim 1, wherein the molecules represented by the product strings are made in parallel in an array of vessels.

33
49. (previously added) The method of claim 1, wherein the molecules represented by the product strings are made by assembly of oligonucleotides.

34
50. (previously added) The method of claim 1, further comprising testing members of the data structure of product strings for a particular property and determining an optimal combination of sequences using multi-variate analysis.

E3
corr.

35   16
51. (previously added) The computer program product of claim 17, wherein the initial character strings of (i) are related.

36   16
52. (previously added) The computer program product of claim 17, wherein the code instructs physical screening of the molecule(s) represented by the product strings for one or more desired properties.

37   16
53. (previously added) The computer program product of claim 17, wherein the code instructs determination of a computationally predicted property for molecules represented by the product strings.

38   16
54. (previously added) The computer program product of claim 17, wherein the molecules represented by the product strings are made in parallel in an array of vessels.

39   16
55. (previously added) The computer program product of claim 17, wherein the molecules represented by the product strings are made by assembly of oligonucleotides.

40   16
56. (previously added) The computer program product of claim 17, wherein the code tests members of the data structure of product strings for a particular property and determines an optimal combination of initial sequences using multi-variate analysis.

41 57. (new)   A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

i) encoding two or more related biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings;

iv) adding the product strings to a data structure to populate a data structure of product strings; and

v) determining whether the product strings have at least a predefined measure of similarity with at least one initial character string; and

vi) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings determined to have greater than the predefined value of sequence identity with at least one initial string.

9